

Journal of Biomolecular Techniques • Volume 32(4); 2021 Dec

Article Watch: December, 2021

Published on: Dec 06, 2021

DOI: 10.7171/3fc1f5fe.b58318ea

License: Copyright © 2021 Association of Biomolecular Resource Facilities. All rights reserved.

ABSTRACT

This column highlights recently published articles that are of interest to the readership of the [Journal of Biomolecular Techniques](#). We encourage ABRF members to forward information on articles they feel are important and useful to Clive Slaughter, MCG-UGA Medical Partnership, 1425 Prince Avenue, Athens GA 30606. Tel; (706) 713-2216; Fax; (706) 713-2221; Email; cslaught@uga.edu or to any member of the editorial board. Article summaries reflect the reviewer's opinions and not necessarily those of the Association.

NUCLEIC ACID SEQUENCING

Chen W, Zhao Y, Chen X, Yang Z, Xu X, Bi Y, Chen V, Li J, Choi H, Ernest B, Tran B, Mehta M, Kumar P, Farmer A, Mir A, Mehra U A, Li J-L, Moos M, Xiao W, Wang C. A multicenter study benchmarking single-cell RNA sequencing technologies using reference samples. *Nature Biotechnology* 39;2021:1103-1114.

Single-cell RNA sequencing (scRNA-seq) is now widely used to investigate of the functional status and abundance of cell populations in tissues both in health and disease. Studies of standardized cell mixtures have demonstrated that the choice of scRNA-seq protocol and data processing method significantly affect the accuracy and interpretation of the results. The present report describes a multicenter study to evaluate the effect of technology platform, sample composition and bioinformatic processing in further detail. Study participants in 4 centers test 2 well-characterized cell lines (breast cancer and B lymphocyte) both individually and in mixtures of varying proportions. The cell lines are expanded separately in the participating centers to add a component of biological variability into the study. The authors compare 4 scRNA-seq platforms; 6 scRNA-seq data pre-processing pipelines, 3 for use with unique molecular identifiers (UMIs) for distinguishing individual cells, and 3 for use without UMIs; 8 normalization methods; and 7 batch-correction algorithms. Uncorrected data show large variation across platforms and centers. Different pre-processing and normalization routines are found to contribute to variation in the final output, but the biggest variation is found to result from batch effects and their correction. The output is assessed in terms of ability to separate dissimilar cell types (clusterability) and ability to group similar cell types together (mixability). Some algorithms perform well on one or other of these criteria, but not both; others perform well on neither, depending on the platform that generated the data and the cell mixtures tested. There

are also instances of overcorrection of batch effects that lead to co-clustering and therefore lack of discrimination between breast cancer and B cells. The authors offer recommendations for optimizing and benchmarking platforms and protocols and selecting methods for specific applications. They indicate that variation across sites and platforms are amenable to correction with appropriately chosen computational methods. The reference samples used in this study are freely available for investigators to calibrate or evaluate existing and newly developed scRNA-seq methodology for themselves.

Deveson I W, Gong B, Lai K, Lococo J S, Richmond T A, Schageman J, Zhang Z, Novoradovskaya N, Willey J C, Jones W, Kusko R, Chen G, Madala B S, Blackburn J, Stevanovski I, Bhandari A, Close D, Conroy J, Hubank M, Marella N, Mieczkowski P A, Qiu F, Sebra R, Stetson D, Sun L, Szankasi P, Tan H, Tang L-Y, Arib H, Best H, Burgher B, Bushel P R, Casey F, Cawley S, Chang C-J, Choi J, Dinis J, Duncan D, Eterovic A K, Feng L, Ghosal A, Giorda K, Glenn S, Happe S, Haseley N, Horvath K, Hung L-Y, Jarosz M, Kushwaha G, Li D, Li Q-Z, Li Z, Liu L-C, Liu Z, Ma C, Mason C E, Megherbi D B, Morrison T, Pabón-Peña C, Pirooznia M, Proszek P Z, Raymond A, Rindler P, Ringler R, Scherer A, Shaknovich R, Shi T, Smith M, Song P, Strahl M, Thodima V J, Tom N, Verma S, Wang J, Wu L, Xiao W, Xu C, Yang M, Zhang G, Zhang S, Zhang Y, Shi L, Tong W, Johann D J, Mercer T R, Xu J, Group S O S W. Evaluating the analytical validity of circulating tumor DNA sequencing assays for precision oncology. *Nature Biotechnology* 39;2021:1115-1128.

DNA of tumor origin and sharing mutations with the tumor cells may be detected in patient serum. The detection of such circulating tumor DNA (ctDNA) has encouraged the hope that its identification and quantification by high-throughput sequencing might be used for molecular stratification and therapeutic monitoring of malignancies in the clinical setting. Clinical deployment of such analyses requires definition of diagnostic limits, assessment of reproducibility, and identification of key experimental variables affecting performance. The present report contributes to validation of the methodology by describing a multi-site, cross-platform evaluation of performance among 5 commercial ctDNA assays. Twelve clinical research facilities test reference samples simulating human ctDNA in serum. Detection of the cell-free DNA from a tumor is challenging because it constitutes a small fraction of the low concentration of total circulating DNA, and because the DNA exists in short fragments. The results of the study nevertheless indicate that ctDNA mutations with variant allele frequency above 0.5% may be detected with adequate sensitivity, specificity and precision by any of the

assay platforms. Below a variant allele proportion of 0.5%, however, performance is not presently acceptable. False positive results (false detection of mutations) is rare compared with false negatives, and can be controlled by the use of unique molecular identifiers. At present levels of performance, clinical deployment may be envisaged for purposes of molecular stratification and profiling in the setting of advanced cancer where DNA from the tumor is at its highest level.

METABOLOMICS

Xue J, Derks R J E, Webb B, Billings E M, Aisporna A, Giera M, Siuzdak G. Single quadrupole multiple fragment ion monitoring quantitative mass spectrometry. *Analytical Chemistry* 93;2021:10879-10889.

Multiple reaction monitoring (MRM) of predetermined precursor ion → product ion transitions in a tandem mass spectrometer system has become standard methodology for quantifying multiple analytes in large numbers of samples. However, Xue *et al.* note that if fragmentation voltages are increased in the source region of the mass spectrometer, fragments are formed in the source region that are generally the same as the ones generated in the collision cell during tandem mass spectrometry, and these in-source fragments could in principle be used for quantification with a single quadrupole mass analyzer lacking a collision cell, provided appropriate precautions are taken to detect coeluting analytes. Indeed, methods of this kind have long been used to quantify small molecules by gas chromatography-mass spectrometry (GC-MS) with electron impact ionization. Such methodology offers potential advantages in sensitivity because ions are not lost in transmission between instrument sectors. There are, of course, also definite advantages of reduced instrument cost. The authors here report results with enhanced in-source fragmentation for metabolite quantification using both mixtures of standards and blood plasma or cell extracts. Using a triple quadrupole platform, they compare dynamic range and precision in parallel experiments, and show comparable or superior performance with enhanced in-source fragmentation. The key advantage of MRM is the selectivity provided by tandem mass filters. Nevertheless, the authors show that matrix effects for 50 metabolites spiked at various concentrations into biological fluids, although compound-specific, are independent of the analytic technique, and accuracy achieved with in-source fragmentation is excellent. The authors also describe a correlated ion monitoring algorithm that autonomously compiles chromatographic data for multiple ions to allow detection of coeluting species based on precursor and fragment ion ratios. Their results indicate that single quadrupoles provide comparable quantification

performance to MRM. The methodology also lends itself to acquisition of pseudo-MS³ spectra for quantifying complex molecules such as oxidized phospholipids or closely related eicosanoids.

MACROMOLECULAR SYNTHESIS & SYNTHETIC BIOLOGY

Huang Y, Knouse K W, Qiu S, Hao W, Padial N M, Vantourout J C, Zheng B, Mercer S E, Lopez-Ogalla J, Narayan R, Olson R E, Blackmond D G, Eastgate M D, Schmidt M A, McDonald I M, Baran P S. A P(V) platform for oligonucleotide synthesis. *Science* 373;2021:1265-1270.

Caruthers' phosphoramidite chemistry for oligonucleotide synthesis has become part of the lore of molecular biotechnology. The coupling reaction in this chemistry initially produces a phosphite linkage with the phosphorus atom in the P(III) oxidation state. The linkage is converted to phosphate, in which the phosphorus atom is in the P(V) oxidation state, by oxidation in a solution of iodine in tetrahydrofuran/pyridine. Attempts to accomplish the coupling directly with P(V) reagents have historically shown such reaction to be comparatively slow. On the other hand, coupling with P(III) reagents to produce unnatural linkages such as phosphorothioate (PS) or phosphorodithioate (PS₂), which are favored for pharmaceutical development, is less facile with P(III) reagents. Huang *et al.* now provide details of a unified suite of P(V)-based reagents and protocols by which unnatural linkages may be incorporated into oligonucleotides at will in commercial synthesizers. Their method departs entirely from the canonical P(III) chemistry. The new methodology permits formation of stereospecific *R*-PS and *S*-PS linkages and their racemic mixtures, as well as PS₂ linkages, and native phosphodiester at desired locations. It also enables incorporation of locked nucleic acid (LNA) as well as DNA sugars to be incorporated into the oligonucleotide backbone at positions of choice. The bases A, T, G & mC are utilized. The coupling scheme in the new chemistry is relatively simple. The phosphorus oxidation step is, of course, omitted, and the cyanoethyl protecting group on the phosphate is eliminated. To address the poor reaction efficiency previously associated with P(V) methods, the authors demonstrate that, with optimized materials and protocols, the methodology produces desired product in high yield. The results with this first iteration of the technology stimulate optimism that the methodology will help in development of therapeutics, where the unnatural linkages are desirable for enhancement of the pharmacokinetic properties and efficacy of the oligonucleotide products.

Reinkemeier C D, Lemke E A. Dual film-like organelles enable spatial separation of orthogonal eukaryotic translation. *Cell* 184;2021:4886-4903.e4821.

A central problem in synthetic biology is how to program new functionality into living cells without interfering with endogenous processes which utilize some of the same components. A approach to this problem is the localization of new functions within membrane-bound or membraneless compartments. Reinkemeier *et al.* explore a related idea: selectively colocalizing components on membrane surfaces. Cells have long used the same approach for localization of signaling pathways. Indeed, analogous processes might have spurred the development of organelles in the course of evolution of eukaryotic cells. The authors of the present paper utilize codon expansion to repurpose a stop codon (the amber codon, TAG) to incorporate a non-canonical amino acid into a protein of interest. An aminoacyl-tRNA synthase that recognizes a cognate, repurposed tRNA is directed to a membrane along with the mRNA of the protein of interest. At that membrane site, the protein incorporates the novel amino acid without misdirecting the stop codon in other mRNAs translated elsewhere. The authors further repurpose the same stop codon to incorporate different non-canonical amino acids into different proteins that are synthesized on different membrane sites - the plasma membrane and the endoplasmic reticulum membrane. The mRNAs are directed to the respective membranes along with appropriate aminoacyl-tRNA synthases. Only the mRNAs and aminoacyl tRNA synthases need be membrane-bound in this methodology. Orthogonal operation of 3 different translational programs (one natural and two engineered) at three spatially distinct sites within the same cell is achieved in this way! This proof-of-concept study indicates a capability to program cells to perform tasks of remarkable complexity by tuning distinct translational outputs at different locations in close proximity within the cell.

Garabedian M V, Wang W, Dabdoub J B, Tong M, Caldwell R MThe a, Benman W, Schuster B S, Deiters A, Good M C. Designer membraneless organelles sequester native factors for control of cell behavior. *Nature Chemical Biology* 17;2021:998-1007.

Garabedian *et al.* provide methodology for constructing membraneless organelles that exist as condensates of engineered, intrinsically disordered proteins. Their purpose is to sequester targeted cell proteins within these organelles, and to re-release the sequestered proteins as desired. The protein scaffold for the organelle is based on the disordered arginine/glycine-rich (RGG) domain of the P granule protein LAF-1. The

authors incorporate 3 RGG domains into the scaffold protein to optimize condensate formation and temperature stability. A high-affinity coiled coil (CC) sequence is added at the *N*-terminus, and cognate CC tags are placed at the C-termini of cellular client proteins to sequester them by binding to the scaffold protein. A high proportion of the scaffold protein is shown to localize to condensates, and partition of client proteins into the condensate phase is also strong. The coiled coil interaction is temperature-sensitive: release is achieved by raising the temperature from 37°C to 42°C. The present experiments are conducted principally with yeast cells, which can tolerate such a temperature jump. For implementation in mammalian cells, an alternative, optical release system based on a photocleavable linker is described. The authors demonstrate the system by arresting the yeast cell cycle by sequestration of Cdc24 at low temperature and release of arrest at high temperature. This cellular engineering methodology therefore provides a means for experimental control of cell function through regulation of the subcellular distribution of selected protein components.

MASS SPECTROMETRY

Baba T, Ryumin P, Duchoslav E, Chen K, Chelur A, Loyd B, Chernushevich I. Dissociation of biomolecules by an intense low-energy electron beam in a high sensitivity time-of-flight mass spectrometer. *Journal of the American Society for Mass Spectrometry* 32;2021:1964-1975.

Beckman J S, Voinov V G, Hare M, Sturgeon D, Vasil'ev Y, Oppenheimer D, Shaw J B, Wu S, Glaskin R, Klein C, Schwarzer C, Stafford G. Improved protein and PTM characterization with a practical electron-based fragmentation on Q-TOF instruments. *Journal of the American Society for Mass Spectrometry* 32;2021:2081-2091.

Two groups describe the addition of electron-activated dissociation (EAD) capability to hybrid quadrupole-time-of-flight (Q-TOF) tandem mass spectrometer systems. Both groups install an EAD cell between the two sectors of an existing Q-TOF instrument – a Sciex X500B in the case of Baba *et al.* and an Agilent 6500 LC/Q-TOF in the case of Beckman *et al.* The ion optic strategies differ between the two groups, but, in both cases, electrons are trapped within the cell at near-zero eV kinetic energies. But very high electron density, which approaches the space-charge limit, maximizes reaction speed. Both groups demonstrate electron capture dissociation (ECD) on their instrument platforms. Baba *et al.* additionally employ tunable electron energy to expand the repertoire of excitation modes to include hot ECD (ECD with enhanced electron kinetic energy) and EIEIO (electron impact excitation of ions from organics, in

which small molecular ions, *e.g.* complex lipids, may be informatively fragmented despite their low charge state). The authors show that in many applications reaction efficiencies and instrument platform sensitivity support data acquisition on the liquid chromatography-mass spectrometry (LC-MS) timescale. EAD provides fragmentation patterns complementary to those of collisional dissociation for a wide range of analytes. The present instrumentation developments are expected to contribute to deployment of EAD methods more widely.

Cai J, Yan Z. Re-examining the impact of minimal scans in liquid chromatography-mass spectrometry analysis. *Journal of the American Society for Mass Spectrometry* 32;2021:2110-2122.

For quantification of analytes by LC-MS, accurate measurement of the area under a chromatographic peak is generally supposed to require 13-20 MS scans. In the present study, this prerequisite is reexamined to ascertain whether a smaller number of scans might suffice. If so, reduction in number of required scans would create openings for higher throughput or greater richness of datasets, at least in non-regulatory assays. On the basis of experiments with a mixture of drugs in the presence and absence of biological matrices, the authors of this paper conclude that 6 scans per analyte are sufficient to achieve high accuracy with contemporary LC and MS systems. They go on to illustrate the advantages of reducing scan numbers in a study of hepatic drug metabolism. The authors argue that the number of data points required adequately to define a peak depends upon the quality of peak detection and peak integration algorithms, and on peak symmetry and separation, all of which may change as instrumentation and methodology improve. Occasional reassessment of conventional prerequisites such as these may therefore prove rewarding.

SEPARATIONS

Plumb R S, McDonald T, Rainville P D, Hill J, Gethings L A, Johnson K A, Wilson I D. High-throughput UHPLC/MS/MS-based metabolic profiling using a vacuum jacketed column. *Analytical Chemistry* 93;2021:10644-10652.

Each new innovation in separation technology highlights previously under-appreciated factors that limit performance: addressing these new factors may yield yet further incremental improvements in separation efficiency. During liquid chromatography (LC) in the ultra-high performance LC (UHPLC) domain, eluent is pumped through columns at pressures in the range 10,000-15,000 psi. Band broadening occurs as a result of viscous heat generated by friction of eluent flowing through the packed column bed.

Temperature gradients arise in both the radial and longitudinal dimensions. Longitudinal heating has previously been addressed by supplying heat to the inlet fitting to maintain a constant temperature along the column length. In the present paper, radial heating is addressed by use of a vacuum-jacketed column and, when performing on-line electrospray LC-MS (LC-ESIMS), it is addressed by providing continuous connecting tubing between the column outlet and the electrospray emitter. The authors document the performance enhancement available with these modifications in a study of metabolites in urine. They perform separation on a 2.1 x 30 mm, 1.7 μ C18 column. Gradient elution is conducted at a flow rate of 1 mL/min. With a 75-s gradient, they achieve an average band width of 0.6 s with a peak tailing factor of 1.13 and peak capacity of 120. With a gradient of 37 s they observe peak widths of ~0.4 s and a peak capacity of 84. Deployment of vacuum jacketing results in an 1.85x increase in peak capacity and a 25% increase in the number of features detected in urine.

PROTEOMICS

Liang Y, Acor H, Mccown M A, Nwosu A J, Boekweg H, Axtell N B, Truong T, Cong Y, Payne S H, Kelly R T. Fully automated sample processing and analysis workflow for low-input proteome profiling. *Analytical Chemistry* 93;2021:1658-1666.

Martin K, Zhang T, Lin T-T, Habowski A N, Zhao R, Tsai C-F, Chrisler W B, Sontag R L, Orton D J, Lu Y-J, Rodland K D, Yang B, Liu T, Smith R D, Qian W-J, Waterman M L, Wiley H S, Shi T. Facile one-pot nanoproteomics for label-free proteome profiling of 50-1000 mammalian cells. *Journal of Proteome Research* 20;2021:4452-4461.

Proteomic analysis of small numbers of cells (<1000) is required for discriminating populations of cells at high levels of spatial resolution. Although the performance of contemporary LC and MS systems is well able to support such studies, attention to sample processing is necessary to avoid prohibitive sample losses resulting from adsorption to surfaces. Liang *et al.* use robotic pipetting with samples housed in 384-well plates to accomplish one-pot sample processing. They employ sample volumes in the low μ L range and correct for evaporation by periodically adding water or buffer to the microwells. They show identification of 1095 proteins from ~130 lymphocytes collected by fluorescence-activated cell sorting (FACS). Martin *et al.*, however, use conventional processing volumes in the range ~50-100 μ L, but avoid transfers between tubes and employ a surfactant, 0.2% β -d-maltoside, to minimize adsorptive

losses. They conduct a one-pot protocol in which all reactions are performed in a single polymerase chain reaction tube, which is also used for LC-MS sample loading. They achieve reliable, label-free quantification of ~1200-2700 proteins from breast cells recovered by FACS and ~1500-2500 proteins from mouse colonic crypt cells. Compatibility with convenient processing volumes is expected to encourage more widespread analysis of low-input samples.

FUNCTIONAL GENOMICS/PROTEOMICS

Parvez S, Herdman C, Beerens M, Chakraborti K, Harmer Z P, Yeh J-R J, Macrae C A, Yost H J, Peterson R T. MIC-Drop: A platform for large-scale in vivo CRISPR screens,. *Science* 373;2021:1146-1151.

In vertebrates, forward genetic screens, in which the gene responsible for a phenotype induced by chemical or insertional mutagenesis is localized and identified, have been invaluable. But the procedure is exceedingly slow and resource-intensive. Reverse genetic screens in vertebrates, conducted by targeted mutagenesis, *e.g.* with CRISPR, are also effective, but are typically severely limited in throughput. The present work shows how to perform reverse genetic screens in zebrafish embryos at high throughput. Microfluidic techniques are used to create 100- μ m droplets, each droplet containing Cas9 and 4 guide RNAs (gRNAs) specific for a particular zebrafish gene, plus a barcode associated with that gene. Droplets targeting hundreds to thousands of different genes are mixed together and injected into single-cell embryos, one droplet per embryo, from a single needle. Embryos are inspected for phenotypes of interest and the perturbed genes are identified by retrieving the barcodes by genome sequencing. This protocol permits large numbers of animals to be treated and avoids the need to separate animals according to the gene disrupted in each. Arguing that heart disease is a premier cause of congenital illness in humans, Parvez *et al.* demonstrate their method by screening zebrafish for genes that cause abnormal cardiac development with high penetrance. Among 188 targeted genes. they identify 1 associated with porphyria, 2 with arrhythmia, and 7 with abnormal cardiac development including defects in ventricular morphogenesis, cardiac looping and formation of the atrioventricular valve. These results show how the new platform can identify genes of putative interest for human health and disease.

Vaninsberghe M, Van Den Berg J, Andersson-Rolf A, Clevers H, Van Oudenaarden A. Single-cell Ribo-seq reveals cell cycle-dependent translational pausing. *Nature* 597;2021:561-565.

The contribution to cell differentiation or pathology made by changes in translation remains poorly charted territory that awaits the development of convenient methodology for measurement of the translation of the products of individual genes at the single cell level. Vaninsberghe *et al.* describe a new single-cell ribosome profiling procedure for this purpose. They sort single live cells into a lysis buffer containing cycloheximide to stabilize the interaction between ribosomes and transcripts and halt translation. Micrococcal nuclease is then used to digest exposed RNA, leaving ribosome-protected sequences. Adaptors containing unique molecular identifiers and priming sites for subsequent cDNA synthesis and indexing PCR are ligated to these footprint sequences, and the products are pooled and size-selected to enrich for those of typical ribosome-protected footprint length. Sequencing the resulting libraries provides a snapshot of translation in progress at the single-cell level. The authors use this methodology to investigate the pausing of translation upon deprivation of particular amino acids, alterations in transcription during mitosis, and ribosome pausing in different sub-populations of mouse enteroendocrine cells that are distinguished by their expression of particular hormone markers.

Altae-Tran H, Kannan S, Demircioglu F E, Oshiro R, Nety S P, Mckay L J, Dlakić M, Inskeep W P, Makarova K S, Macrae R K, Koonin E V, Zhang F. The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. *Science* 374;2021:57-65.

With the initial intention to study the evolutionary origin of the RNA-guided bacterial nuclease Cas9 and the function of its ancestors, Altae-Tran *et al.* unexpectedly discover a huge number of programmable nucleases that represent a resource of potentially diverse reagents for future biotechnological applications. Cas9 itself is a nuclease of bacterial and archaeal cells that functions as an antiviral enzyme. It specifically degrades the DNA of viruses previously encountered by its bacterial/archaeal host. Its specificity is guided by complementary RNA sequences encoded in a genomic CRISPR array located in close proximity to the nuclease gene. The evolutionary ancestor of Cas9 is believed to have arisen within a family of putative nucleases called *IscB*, which are encoded by the IS200/IS605 family of transposons. Altae-Tran *et al.* identify CRISPR arrays close to some *IscB* genes, suggesting that some, possibly all, *IscB* proteins might also be RNA-programmable nucleases, whether they are encoded by genes close to CRISPR arrays or not. The authors subsequently find sequences encoding previously unknown RNAs, which they name ω RNAs, close to *IscB* genes. They further show that these ω RNAs can act as guide RNAs for *IscB* proteins in the cleavage of double-stranded DNA. The function of *IscB* nucleases

remains unclear, but it might be involved in transposition or transposon maintenance. Having demonstrated that one family of transposon-encoded nucleases can act as RNA-guided nucleases, Altae-Tran *et al.* go on to show that another such family called TnpB, which is believed to have produced the ancestor of Cas12, is also RNA-guided. *TnpB* is a huge prokaryotic gene family. This work therefore reveals an extensive resource for the development of biotechnology tools in the future.

MACROMOLECULAR CHARACTERIZATION

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, Bridgland A, Meyer C, Kohl S a A, Ballard A J, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior A W, Kavukcuoglu K, Kohli P, Hassabis D. Highly accurate protein structure prediction with AlphaFold. *Nature* 596;2021:583-589.

Baek M, Dimaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee G R, Wang J, Cong Q, Kinch L N, Schaeffer R D, Millán C, Park H, Adams C, Glassman C R, Degiovanni A, Pereira J H, Rodrigues A V, Van Dijk A A, Ebrecht A C, Opperman D J, Sagmeister T, Buhlheller C, Pavkov-Keller T, Rathinaswamy M K, Dalwadi U, Yip C K, Burke J E, Garcia K C, Grishin N V, Adams P D, Read R J, Baker D. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373;2021:871-876.

Publications by two groups describe the results of major improvements in the prediction of three-dimensional structure from protein amino acid sequence. Jumper *et al.* describe the capabilities of the neural network, AlphaFold 2, a product of DeepMind, a London-based artificial intelligence company owned by Google. As the bases for prediction, AlphaFold2 combines bioinformatic and biophysical information. It uses alignments with multiple homologous amino acid sequences, along with physical and geometric rules governing protein structure. AlphaFold2 achieves unprecedented spatial accuracy: an all-atom accuracy of 1.5 Å r.m.s.d.₉₅ (95% confidence interval = 1.2-1.6 Å), compared to the best prior method's 3.5 Å r.m.s.d.₉₅ (95% confidence interval = 3.1-4.2 Å). Such enhanced performance emulates that of experimental structures. The method is scalable to very long polypeptides and very large numbers of proteins. It also provides estimates of reliability for each residue's position. AlphaFold2 may be used by academic researchers upon request, and its predicted structures are also freely available. Inspired by the first announcement of

this method at the Critical Assessment of Structure Prediction meeting in 2020, Baek *et al.* report that they were motivated to create their own network architecture along similar lines, and additionally to make the code for their network, named RoseTTAFold, generally available to the scientific community. They report accuracies approaching those of Jumper *et al.* with their network. Baek *et al.* show how their predictive models enable the solution of structures using X-ray crystallographic data from molecular replacement and from electron density maps obtained by cryo-electron microscopy. They use structural predictions to help explain how mutations in specific proteins may cause disease. Finally, they show that their network can build models of protein-protein interfaces based upon amino acid sequence alone, bypassing the conventional process of building models of the individual molecules and then conducting rigid-body docking. The development of these capabilities for structure prediction are widely recognized to be of seminal importance in molecular biology.

Townshend R J L, Eismann S, Watkins A M, Rangan R, Karelina M, Das R, Dror R O. Geometric deep learning of RNA structure. *Science* 373;2021:1047-1051.

Structure prediction for RNA is less well developed than that for proteins. Base pairing can be predicted accurately, but modeling of higher order structures remains uncertain, in part because the number of available experimental structures is small, and in part because sequence coevolution provides less information about tertiary contacts than it does in proteins. Townshend *et al.* describe a neural network called Atomic Rotationally Equivalent Scorer (ARES) that significantly improves prediction quality. As inputs, it uses physical and geometric rules governing structure, but does not incorporate any RNA-specific information (*e.g.* assumptions about the relevancy of structural features such as double helices, hydrogen bonds or sequences of related RNAs). The training set consists of a sparse 18 structures solved prior to 2007. ARES nevertheless predicts structures to an average of ~ 12 Å r.m.s.d., whereas previous methods achieve only ~ 15 -20 Å r.m.s.d. Interestingly, although the RNAs in the training set are only 17-47 nucleotides long, ARES predicts the structures of much longer RNAs: 27-188 nucleotides for RNAs in the benchmark set and 112-230 nucleotides in the challenge set. These results suggest that substantial acceleration in the area of RNA structural studies is imminent.

IMAGING

Silvestri L, Müllenbroich M C, Costantini I, Di Giovanna A P, Mazzamuto G, Franceschini A, Kutra D, Kreshuk A, Checcucci C, Toresano L O, Frasconi P,

Sacconi L, Pavone F S. Universal autofocus for quantitative volumetric microscopy of whole mouse brains. *Nature Methods* 18;2021:953-958.

In the deployment of light-sheet microscopy for imaging across large volumes at high resolution, *e.g.* in the subcellular imaging of entire, clarified mouse brains stained with a fluorescent dye, defocusing is responsible for most image degradation. This results from loss of coincidence between the light sheet and the focal plane of the detection objective. Silvestri *et al.* describe a method for real-time, image-based focus stabilization that is compatible with light-sheet microscopy and does not depend on image content. It derives from the principle of phase detection. Rays passing through distinct portions of a split objective pupil intersect the image plane at different lateral positions when the object is defocused. An auxiliary camera collects the second ray bundle. When the focus varies, lateral displacement of the two images occurs such that the displacement (the 'phase') provides a measure of the focal state of the microscope that may be used for real-time feedback focus stabilization. The authors apply this principle in light-sheet microscopy with various clearing and staining methods. They also demonstrate its use for alternate applications, including *in vivo* fluorescence imaging and tracking of moving organisms.

DRUG DEVELOPMENT

Segel M, Lash B, Song J, Ladha A, Liu C C, Jin X, Mekhedov S L, Macrae R K, Koonin E V, Zhang F. Mammalian retrovirus-like protein PEG10 packages its own mRNA and can be pseudotyped for mRNA delivery. *Science* 373;2021:882-889.

This work represents a proof-of-principle demonstration of a programmable vehicle for delivery of RNA to target cells that provides a potentially advantageous alternative to lipid nanoparticles or viruses such as adenovirus. The components of the delivery vehicle are entirely endogenous to mammals. This feature encourages the hope that the vehicle will prove to be non-immunogenic. A substantial portion of the human genome (>8%) originates from retroviruses or retrotransposons. They may be passed from cell to cell by encapsulation of the RNA genome in a capsid or vesicle. A core structural gene, *gag*, typically encodes RNA-binding matrix and capsid proteins, and the genome is flanked by terminal repeat elements. Interestingly, some retroviral genes are known to have been co-opted to perform indispensable physiological functions for the host. The authors identify a *gag* homolog in mice and humans, *Peg10*, that preferentially binds its own mRNA and facilitates transfer of this mRNA between cells in virus-like particles. They then show that the flanking 5' and 3' untranslated

regions of *Peg10* can be engrafted onto an mRNA cargo to mediate the packaging, secretion and functional delivery of the cargo to recipient cells in the same way. Future development of this system is hoped to provide a new platform for gene transfer of use for experimental or therapeutic applications.

Abbasov M E, Kavanagh M E, Ichu T-A, Lazear M R, Tao Y, Crowley V M, Am Ende C W, Hacker S M, Ho J, Dix M M, Suci R, Hayward M M, Kiessling L L, Cravatt B F. A proteome-wide atlas of lysine-reactive chemistry. *Nature Chemistry* 13;2021:1081-1092.

The search for ligands of proteins that may be developed into drugs has traditionally been driven by high-throughput screening of large libraries of compounds, but this approach is increasingly supplemented by proteome-wide screening to identify the sites at which chemically reactive probes bind to proteins as detected by quantitative mass spectrometry. The present study represents a large expansion of the scope of the latter approach in which the reactivity and chemoselectivity of diverse electrophiles capable of interacting with lysine residues is cataloged on a proteome-wide scale. The authors construct a library of ~180 such aminophilic molecules with 30 different types of chemical reactivity (chemotypes). Most of them comply with the Lipinski rule-of-five values in order to favor the discovery of drug-like lead compounds. These aminophiles are incubated over a range of concentrations with extracts of a lymphoma cell line and a breast cancer cell line, which have rather distinct proteomes. Dimethyl sulfoxide serves as a control. The extracts are then incubated with a broadly reactive tracer aminophile that may be conjugated to isotopically distinct azide-biotin tags (heavy and light) by copper-catalyzed azide-alkyne cycloaddition (CuAAC). The protein pools are combined, enriched by streptavidin, proteolytically digested, and the peptides identified by LC-MS to determine which lysine residues are liganded by members of the aminophile library as indicated by their ability to compete for binding with the indicator electrophile. In this way, the authors identify 818 lysines on 581 proteins that interact with 1 or more of the aminophiles. Some aminophiles have broad reactivity, some highly specific. The lysines are present on proteins with diverse structural and functional characteristics. Follow-up experiments indicate instances in which site-specific binding modifies protein function. The data provide a resource for the discovery of new therapeutics, including ones that affect hitherto difficult to explore functions such as RNA-protein interactions. The study also provides methodology for future high-throughput chemical probe development.